Journal of Agricultural Machinery

Homepage: https://jame.um.ac.ir

Research Article

Vol. 12, No. 2, Summer 2022, p. 159-167

Spectral Feature Selection from the Hyperspectral Dataset to Identify Pistachio Leaves Infected by Psylla

A. Moghimi[1], A. Sazgarnia[2], M. H. Aghkhani[3*]

1- Ph.D. student of Department of Biosystems Engineering, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran
2- Professor, Department of Medical Physics, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran
3- Professor, Department of Biosystems Engineering, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran
(*- Corresponding Author Email: aghkhani@um.ac.ir)

Received: 22-07-2019
Revised: 24-08-2020
Accepted: 02-09-2020

How to cite this article:
Moghimi, A., A. Sazgarnia, and M. H. Aghkhani. 2022. Spectral Feature Selection from the Hyperspectral Dataset to Identify Pistachio Leaves Infected by Psylla. Journal of Agricultural Machinery 12 (2): 159-167. (In Persian). DOI: 10.22067/jam.v12i2.82089.

Introduction

Pistachio production has been adversely affected by Psylla, which is a devastating insect. The primary goal of this study was to select sensitive spectral bands to distinguish pistachio leaves infected by Psylla from healthy leaves. Diagnosis of psylla disease before the onset of visual cues is crucial for making decisions about topical garden management. Since it is not possible to diagnose psylla disease even after the onset of symptoms with the help of color images by drones, hyperspectral and multispectral sensors are needed. The main purpose of this study was to extract spectral bands suitable for distinguishing healthy leaves from psylla leaves. For this purpose, in this paper, a new method for selecting sensitive spectral properties from hyperspectral data with the high spectral resolution is presented. The intelligent selection of sensitive bands is a convenient way to build multispectral sensors for a specific application (in this article, the diagnosis of psylla leaves). Knowledge of disease-sensitive wavelengths can also help researchers analyze multispectral and hyperspectral aerial images captured by satellites or drones.

Materials and Methods

A total number of 160 healthy and diseased leaves were scanned in 64 spectral bands between 400-1100 nm with 10 nm spectral resolution. A random forest algorithm was used to identify the importance of features in classifying the dataset into diseased and healthy leaves. After computing the importance of the features, a clustering algorithm was developed to cluster the most important features into six clusters such that the center of clusters was 50 nm apart. To transfer the hyperspectral dataset into a multispectral dataset, the reflectance was averaged in spectral bands within ±15 nm of each cluster center and achieved six broad multispectral bands. Afterwards a support vector machine algorithm was utilized to classify the diseased and healthy leaves using both hyperspectral and multispectral datasets.

Results and Discussion

The center of clusters were 468 nm, 598 nm, 710 nm, 791 nm, 858 nm, and 1023 nm, which were calculated by taking the average of all the members assigned to the individual clusters. These are the most informative spectral bands to distinguish the pistachio leaves infected by Psylla from the healthy leaves. The F1-score was 90.91 when the hyperspectral dataset (all bands) was used, while the F1-score was 88.69 for the multispectral dataset. The subtle difference between the F1-scores indicates that the proposed pipeline in this study was able to select appropriately the sensitive bands while retaining all relevant information.

Conclusion

The importance of spectral bands in the visible and near-infrared region (between 400 and 1100 nm) was obtained to identify pistachio tree leaves infected with psylla disease. Based on the importance of spectral properties and using a clustering algorithm, six wavelengths were obtained as the best wavelengths for classifying healthy and diseased pistachio leaves. Then, by averaging the wavelengths at a distance of 15 nm from these six centers, the hyperspectral data (64 bands) became multispectral (6 bands). Since the correlation between the wavelengths in the near-infrared region was very high (more than 95%), out of the three selected wavelengths in the near-infrared region (710, 791, and 1023), only the 710-nm wavelength, which was closer to the visible region, was selected. The results of classification of infected and diseased leaves using hyperspectral and multispectral data showed that the degree of classification accuracy decreases by about 2% and if only 4 bands are used, the degree of accuracy decreases by about 3%.

The results of this study revealed that the proposed framework could be used for selecting the most informative spectral bands and accordingly develop custom-designed multispectral sensors for disease detection in pistachio. In addition, we could reduce the dimensionality of the hyperspectral datasets and avoid the issues related to the curse of dimensionalitylity.

Keywords: Classification, Feature selection, Hyperspectral data, Multispectral data, Pistachio, Psylla, Random forest, Spectroscopy

نشریه ماشین‌های کشاورزی
https://jame.um.ac.ir

مقاله پژوهشی
جلد ۱۲، شماره ۲، تابستان ۱۴۰۱، ص ۱۶۷-۱۵۹

انتخاب طول‌موج‌های طیفی حساس از میان داده‌های فراطیفی جهت تشخیص برگ‌های درخت پسته آلوده به پسیل

علی مقیمی[1]، آمنه سازگارنیا[2]، محمد حسین آق خانی[3]*

تاریخ دریافت: ۱۳۹۸/۰۴/۳۱
تاریخ پذیرش: ۱۳۹۹/۰۶/۱۲

چکیده

در سال‌های اخیر تولید پسته توسط آفتی به نام پسیل تهدید شده است. هدف از انجام این تحقیق انتخاب باندهای طیفی مناسب جهت تشخیص برگ‌های آلوده به آفت پسیل می‌باشد. برای این منظور از ۱۶۰ برگ سالم و ۱۶۰ برگ بیمار در ۶۴ باند طیفی در ناحیه بین ۴۰۰ تا ۱۱۰۰ نانومتر با قدرت تفکیک طیفی حدوداً ۱۰ نانومتر طیف‌نگاری انجام شد. از روش جنگل‌های تصادفی برای مرتب‌سازی طول‌موج‌ها بر اساس قدرت تفکیک بین برگ‌های سالم و بیمار استفاده شد. پس از شناسایی و مرتب‌سازی طول‌موج‌ها بر اساس میزان اهمیت، طول‌موج‌های دارای بیشترین اهمیت توسط یک الگوریتم خوشه‌بندی در شش خوشه تقسیم‌بندی شدند به‌طوری‌که میانگین طول‌موج‌های هر خوشه‌ها حداقل ۵۰ نانومتر از یکدیگر فاصله داشته باشند. میانگین طول‌موج‌های قرار گرفته در شش خوشه به‌ترتیب اهمیت عبارت بود از: ۷۱۰، ۷۹۱، ۸۵۸، ۵۹۸، ۴۶۸، و ۱۰۲۳ نانومتر. برای تبدیل داده‌های فراطیفی به چندطیفی، از بازتابش‌های به‌دست آمده در فاصله ±۱۵ نانومتری این مراکز میانگین‌گیری انجام گرفت و داده‌های بازتابش به‌دست آمده از دیگر طیف‌ها حذف شد. الگوریتم ماشین بردار پشتیبان برای طبقه‌بندی برگ‌های سالم و بیمار با استفاده از داده‌های فراطیفی و چند طیفی به‌دست آمده در این پژوهش به‌کار گرفته شد. دقت طبقه‌بندی با استفاده از تمام ۶۴ طول‌موج (داده‌های فراطیفی) ۹۰/۹۱ درصد و با استفاده از ۶ طول‌موج (داده‌های چندطیفی) ۸۸/۶۹ درصد بود. اختلاف بسیار کم (حدود ۲ درصد) در میزان دقت طبقه‌بندی نشان‌دهنده صحت شیوه ارائه شده در این پژوهش برای کاهش ابعاد داده‌های فراطیفی می‌باشد. علاوه بر کاهش ابعاد داده، تعیین باندهای طیفی مناسب از میان داده‌های فراطیفی گامی موثر در طراحی و ساخت حسگری چندطیفی جهت تشخیص بیماری گیاهان می‌باشد.

واژه‌های کلیدی: انتخاب ویژگی، پسته، جنگل‌های تصادفی، طیف‌سنجی، فراطیفی

مقدمه

به‌طور کلی پایش وضعیت سلامت و تشخیص بیماری گیاهان و درختان برای کشاورزی پایدار امری حیاتی می‌باشد. اطلاعات اولیه از سلامت درختان و تشخیص بیماری آن‌ها می‌تواند کنترل بیماری‌ها را تسهیل نماید و در نهایت بهره‌وری را افزایش دهد که این مسئله با به‌کارگیری استراتژی‌های مدیریتی مناسب مانند مدیریت استفاده از آفت‌کش‌ها، قارچ‌کش‌ها و مواد شیمیایی مخصوص بیماری محقق می‌گردد.

یکی از محصولات مهم در حوزه کشاورزی ایران پسته می‌باشد که اهمیت بسیار بالایی در بخش صادرات غیرنفتی ایران دارد (Mohammadigol *et al.*, 2013). عاملی که هر ساله باعث به‌وجود آمدن خسارت‌های مالی فراوانی در باغات پسته ایران می‌شود آفت پسیل می‌باشد. علاوه بر تبعات اقتصادی این آفت، بدون تردید سم‌پاشی‌های بی‌رویه با سموم خطرناک، فجایع زیست‌محیطی و بهداشتی متعددی در پی خواهد داشت. در حال حاضر، پایش وضعیت سلامت درختان پسته بیشتر به بازرسی چشمی و میدانی توسط کشاورزان انجام می‌گیرد. به وضوح این روش نیاز به نیروی کارگری و هزینه بالایی دارد و همچنین یک روش بسیار زمان بر و کم دقت می‌باشد.

در سال‌های اخیر با توجه به دسترسی به پلتفرم‌های هوایی و

۱- دانش‌آموخته دکترای گروه مهندسی بیوسیستم، دانشگاه فردوسی مشهد، مشهد، ایران
۲- استاد، گروه فیزیک پزشکی، دانشگاه علوم پزشکی مشهد، مشهد، ایران
۳- استاد، گروه مهندسی بیوسیستم، دانشگاه فردوسی مشهد، مشهد، ایران
(*- نویسنده مسئول: Email: aghkhani@um.ac.ir)
DOI: 10.22067/jam.v12i2.82089

زمینی خودکار، استفاده از سنسورهای چندطیفی و فراطیفی به‌عنوان یک روش سریع و غیر مخرب برای ارزیابی سلامت درختان مطرح شده است. این سنسورها در ناحیه مرئی و فروسرخ نزدیک عمل می‌کنند. طیف‌سنجی در ناحیه مرئی و فروسرخ می‌تواند به‌عنوان یک روش موثر برای ارزیابی خصوصیات کیفی درونی میوه‌ها و سبزی‌ها (Moghimi *et al.*, 2011)، تشخیص تنش‌های زیستی (Susič *et al.*, 2018) و غیر زیستی (Moghimi *et al.*, 2018) گیاهان استفاده گردد حتی قبل از این‌که نشانه‌های بصری به‌وجود آیند. هنگامی که گیاه دچار تنش می‌شود باز تابش گیاه در ناحیه‌های مرئی و فروسرخ نزدیک تغییر می‌کند. آسیب در ساختار برگ، محتوی آبی درون سلول، کاهش میزان کلروفیل و تغییرات رنگدانه از فعل و انفعال بیماری در گیاه منتج می‌شود و این مسئله روی بازتابش گیاه تاثیر می‌گذارد (Mahlein *et al.*, 2012; Wahabzada *et al.*, 2016).

در اینجا به بررسی چند نمونه از پژوهش‌های انجام گرفته در زمینه استفاده از سنسورهای فراطیفی و چندطیفی در کشاورزی می‌پردازیم. جهت تشخیص آلودگی آفلاتوکسین در پسته از تصویربرداری فراطیفی و برای طبقه‌بندی از الگوریتم‌های یادگیری ماشین استفاده شد (Kheiralipour *et al.*, 2016). دقت به‌دست آمده برای جداسازی پسته‌های آلوده از سالم به میزان ۹۱/۷ درصد بود.

برای تشخیص بیماری کپک سطحی (Powdery mildew) در جو از تصویربرداری فراطیفی و برای تحلیل تصاویر از ماشین بردار پشتیبان استفاده شد (Thomas *et al.*, 2018). در این پژوهش میزان شدت بیماری به محض بروز اولین نشانه‌های بصری توسط تصویربرداری فراطیفی قابل تشخیص بود.

در تحقیقی بر میزان اهمیت کاهش ابعاد داده‌های فراطیفی جهت تشخیص بیماری پوسیدگی زغالی پرداخته شد (Nagasubramanian *et al.*, 2018). در این پژوهش طول‌موج‌های حساس که قادر به تشخیص بیماری به‌صورت موثر بودند انتخاب شدند. از میان ۲۴۰ طول‌موج دوربین فراطیفی، فقط تعداد شش طول‌موج (۴۷۵، ۵۱۶، ۵۴۸، ۶۵۲، ۷۲۰، و ۹۱۵ نانومتر) برای رسیدن به دقت طبقه‌بندی ۹۷ درصد لازم بود که بیانگر اهمیت انتخاب طول‌موج‌های حساس می‌باشد.

در تحقیقی مشابه از دوربین فراطیفی و الگوریتم‌های یادگیری ماشین برای انتخاب طول‌موج‌های حساس جهت تشخیص بیماری FHB[1] در گندم استفاده شد (Moghimi *et al.*, 2019). نتایج این پژوهش نشان داد که تعداد ۵ طول‌موج به‌ترتیب اهمیت شامل ۷۶۶، ۸۶۸، ۶۹۶، ۵۹۱، و ۵۴۰ نانومتر برای تشخیص بیماری کافی می‌باشد. در پژوهشی دیگر از دوربین فراطیفی جهت تشخیص خوشه‌های هفت واریته مختلف گندم آلوده به بیماری FHB استفاده شد (Alisaac *et al.*, 2018). تصویربرداری در ناحیه مرئی و فروسرخ (۴۰۰ تا ۲۵۰۰ نانومتر) انجام شد و نتایج نشان داد که همبستگی بالایی بین شدت بیماری و چهار ناحیه از طول‌موج‌ها (۴۳۰ تا ۵۲۵ نانومتر، ۵۶۰ تا ۷۱۰ نانومتر، ۷۴۰ تا ۸۱۰ نانومتر و ۱۱۱۵ تا ۲۵۰۰ نانومتر) وجود دارد.

به‌طور کلی دوربین‌های فراطیفی به علت رزولوشن طیفی بالا ابزار بسیار مناسبی برای بررسی بازتابش برگ‌های دچار تنش و مقایسه آن با برگ‌های سالم می‌باشند. به هر حال عواملی مانند قیمت بالا، همبستگی بالا میان طیف‌ها، پیچیدگی تحلیل و تفسیر داده‌ها و مشکلات مربوط به نوع تصویربرداری (Line scanning) این دوربین‌ها عواملی هستند که استفاده از این نوع سنسورها را محدود به انجام مطالعات پژوهشی کرده است. از طرفی دوربین‌های چندطیفی ممکن است قادر به تشخیص تنش گیاه نباشند به علت رزولوشن طیفی پایین و همچنین این نکته که آفت و بیماری‌های گوناگون به صورت‌های مختلفی بر نحوه بازتابش برگ‌ها اثر می‌گذارند. لذا یک دوربین چند طیفی که قادر به سنجش در طیف‌های خاص و از پیش تعریف شده است ممکن است قادر به تشخیص تنش‌های محدودی باشد.

تشخیص بیماری پسیل قبل از ظهور نشانه‌های بصری جهت اتخاذ تصمیمات در راستای مدیریت موضعی باغ بسیار مهم می‌باشد. از آنجا که امکان تشخیص بیماری پسیل حتی پس از ظهور علائم بیماری با کمک تصاویر رنگی گرفته شد توسط پرنده‌های بدون سرنشین وجود ندارد نیاز به سنسورهای فراطیفی و چندطیفی می‌باشد. هدف عمده این مطالعه استخراج باندهای طیفی مناسب برای تشخیص برگ سالم از برگ‌های دچار به آفت پسیل می‌باشد. برای این منظور در این مقاله یک روش جدید برای انتخاب ویژگی‌های طیفی حساس از میان داده‌های فراطیفی با قدرت تفکیک طیفی بالا ارائه می‌شود. انتخاب هوشمندانه باندهای حساس یک راه‌حل مناسب برای ساخت سنسورهای چندطیفی برای یک کاربرد خاص (در این مقاله تشخیص برگ‌های دچار به بیماری پسیل) می‌باشد. همچنین اطلاع از طول‌موج‌های حساس به بیماری می‌توانند به محققین در تحلیل تصاویر هوایی چندطیفی و فراطیفی اخذ شده به‌وسیله ماهواره و یا پرنده‌های بدون سرنیشین کمک شایانی نماید.

مواد و روش‌ها

آماده‌سازی برگ‌ها

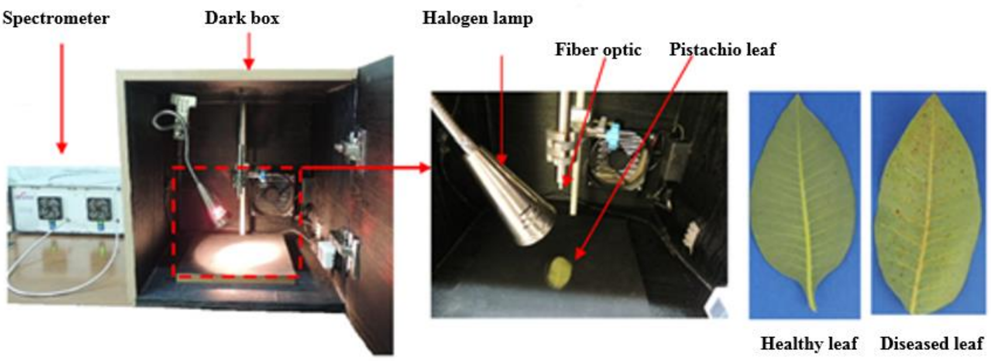
برگ‌ها از یک باغ پسته به مساحت ۲۵ هکتار واقع در روستای حسن آباد کیلومتر ۴۰ جاده مشهد-سرخس با طول جغرافیایی

1- Fusarium head blight

E″۶۰°۰۱′۱۲ و عرض جغرافیایی N″۳۶°۰۶′۳۰ برداشت شده است. انتخاب برگ‌های سالم و بیمار به‌صورت بصری و بر اساس علائم آفت پسیل (وجود لکه و عسلک روی برگ) صورت گرفت (شکل ۱). هر یک از برگ‌های سالم و بیمار پس از آن‌که از درخت جدا شدند در داخل پلاستیک قرار گرفتند. در طول زمان انتقال به آزمایشگاه تمام پلاستیک‌ها در داخل جعبه حاوی یخ نگهداری شدند و بلافاصله در همان روز به آزمایشگاه منتقل گشتند. در مجموع از ۳۲۰ برگ (شامل ۱۶۰ برگ سالم و ۱۶۰ برگ آفت زده) طیف‌نگاری انجام گرفت.

طیف‌سنجی ابرطیفی

طیف‌سنجی در همان روز برداشت برگ‌ها توسط اسپکترومتر آزمایشگاهی (AvaSpec-2048 Dual Thermo-Electric Cooled Fiber Optic) انجام شد. دامنه اسکن دستگاه بین ۴۰۰ نانومتر تا ۱۱۰۰ نانومتر بود و در مجموع ۶۴ باند با فاصله حدودا ۱۰ نانومتر برای هر برگ اسکن شد. برای انجام طیف‌نگاری به‌صورت یکسان در طول انجام آزمایش، طیف‌نگاری از قسمت میانی و اطراف رگبرگ اصلی برگ‌ها انجام شد. از یک تفلون سفید به‌عنوان مرجع جهت نرمالیزه کردن بازتابش برگ استفاده شد. هر ده دقیقه یکبار از تفلون سفید طیف‌نگاری شد زیرا میزان تابش منبع نوری مورد استفاده ممکن است در طول زمان تغییر کند. منبع نوری لامپ هالوژنی اسرام (۵۰ وات) بود که قادر به ساطع کردن نور در طول‌موج‌های مرئی و فروسرخ نور می‌باشد. نور بازتاب شده از سطح برگ‌ها با استفاده از فیبر نوری (standard reflection probe with SMA connector) به سنسور دستگاه اسپکترومتر هدایت شد. به‌منظور جلوگیری از تداخل بین نورهای محیطی و نور بازتاب شده، تمام آزمایش‌ها در فضای کاملاً تاریک انجام گرفت. برای این منظور اتاقک کوچکی با دیواره‌های تیره برای جلوگیری از بازتابش نور تهیه شد و طیف‌نگاری در داخل آن انجام گرفت. شکل ۱ نحوه طیف‌نگاری از برگ‌های سالم و بیمار را نشان می‌دهد.



شکل ۱- طیف‌نگاری برگ‌های پسته با استفاده از اسپکترومتر و نمونه برگ‌های سالم و بیمار

Fig. 1. Scanning the healthy and diseased leaves of pistachio using a spectrometer

پیش‌پردازش داده‌ها

ابتدا میزان جریان تاریک (Dark current) اسپکرومتر اندازه‌گیری و سپس از طول‌موج‌های به‌دست آمده از برگ و تفلون کم شد. در گام بعدی، برای از بین بردن اثر تغییرات شدت نور ساطع شده از منبع نوری در طول آزمایش، طیف‌های خام برگ‌ها به طیف به‌دست آمده از تفلون سفید تقسیم شد. با انجام این پیش‌پردازش بازتابش نسبی برگ‌ها که بین صفر و یک می‌باشد به‌دست آمد (Moghimi *et al.*, 2020). همچنین داده‌ها نرمال‌سازی (با میانگین صفر و واریانس یک) شدند.

انتخاب باندهای حساس

از الگوریتم جنگل‌های تصادفی استفاده شد که ترکیبی از چندین درخت تصمیم‌گیری می‌باشد (Breiman, 2001). الگوریتم جنگل‌های تصادفی در این تحقیق دارای ۱۰۰۰ درخت بود. هر درخت با استفاده از مجموعه‌ای از داده‌ها که به‌صورت تصادفی از میان داده‌های آموزش انتخاب شدند آموزش دیده شد. همچنین برای هر درخت مجموعه‌ای از ویژگی‌ها به‌صورت تصادفی از میان کل ویژگی‌ها انتخاب شد. تعداد ویژگی‌های انتخاب شده برای هر درخت در این تحقیق برابر با مقدار جذر تعداد کل ویژگی‌ها بود (Hastie *et al.*, 2009).

از معیار آنتروپی برای انتخاب ویژگی جهت آزمون در هر گره از درخت استفاده گردید. ویژگی که دارای بیشترین قدرت جداسازی نسبت به سایر ویژگی‌ها بود برای آن گره انتخاب شد تا داده‌ها به دسته‌های کوچک‌تر و خالص‌تر در هر گره تقسیم شوند. بر اساس این نکته که چه مقدار ویژگی انتخاب شده قادر به تفکیک داده‌ها می‌باشد اهمیت آن ویژگی برای آن درخت محاسبه و سپس این مقدار با مقادیر به‌دست آمده از دیگر درخت‌ها جمع و به تعداد کل درختان

تقسیم شد تا اهمیت آن ویژگی برای کل درخت‌ها میانگین گرفته شده باشد.

از نرم‌افزار Python و ابزار scikit-learn برای آموزش جنگل‌های تصادفی و انتخاب ویژگی استفاده شد.

خوشه‌بندی ویژگی‌های طیفی (طول‌موج‌ها)

پس از آن‌که ویژگی‌ها به‌ترتیب اهمیت توسط جنگل‌های تصادفی مرتب شدند در مرحله بعدی ویژگی‌های مهم در شش کلاس خوشه‌بندی شدند (Moghimi *et al.*, 2018). از آنجا که بیشتر دوربین‌های چندطیفی حداکثر دارای شش کانال می‌باشند از یک الگوریتم برای خوشه‌بندی ویژگی‌ها (که به‌ترتیب اهمیت مرتب شدند) در شش خوشه استفاده شد. این الگوریتم خوشه‌بندی ویژگی‌ها را به‌صورت ترتیبی از اولین ویژگی مهم شروع می‌کند. بدین صورت که ویژگی به خوشه‌ای تعلق می‌گیرد که اختلاف بین ویژگی و مرکز آن خوشه کمتر از ۵۰ نانومتر باشد. پس از هر بار خوشه‌بندی هر یک از ویژگی‌ها به یکی از خوشه‌ها، مرکز آن خوشه با میانگین‌گیری از ویژگی‌های آن خوشه به‌روز می‌شود و در تکرارهای بعدی از مراکز جدید خوشه‌ها برای اندازه‌گیری اختلاف بین ویژگی‌ها و مراکز خوشه‌ها استفاده می‌شود.

روند کار به این صورت است که اولین ویژگی در اولین خوشه قرار گرفته و مرکز خوشه به‌روز می‌شود. پس از پایان تکرار اول الگوریتم، فقط مرکز خوشه اول به روز می‌شود. سپس الگوریتم ویژگی بعدی را ارزیابی می‌کند. اگر این ویژگی در فاصله ۵۰ نانومتری مرکز خوشه اول باشد آن‌را در همان خوشه اول قرار داده در غیر این‌صورت به خوشه دوم منتقل می‌شود و مراکز خوشه‌ها با میانگین‌گیری از ویژگی‌های قرار گرفته در آن خوشه به‌روز می‌گردد. روند کار وقتی پایان می‌یابد که الگوریتم نتواند ویژگی بعدی را در یکی از شش خوشه قرار دهد زیرا اختلاف بین این ویژگی و هر یک از مراکز خوشه‌ها بیشتر از ۵۰ نانومتر است.

میانگین‌گیری از باندها و کاهش بعد و تبدیل به چندطیفی

پس از خوشه‌بندی طول‌موج‌ها در شش خوشه، از مراکز خوشه‌ها برای تبدیل داده‌های فراطیفی به چندطیفی استفاده شد به این صورت که از بازتابش‌های به‌دست آمده در فاصله ±۱۵ نانومتری این مراکز میانگین‌گیری انجام گرفت و بازتابش در دیگر طیف‌ها حذف شدند. انتخاب بازه ±۱۵ نانومتری جهت میانگین‌گیری بدان علت بود که برای ساخت دوربین‌های چندطیفی بتوان از فیلترهایی با دامنه عبور ۳۰ نانومتر استفاده نمود.

طبقه‌بندی برگ‌های بیمار و سالم

پس از تبدیل داده‌های فراطیفی و چندطیفی از مدل ماشین بردار پشتیبان[1] برای طبقه‌بندی برگ‌های بیمار از سالم استفاده گردید. در این تحقیق از روش اعتبارسنجی متقابل (Cross-validation) برای ارزیابی میزان قابلیت تعمیم مدل و همچنین میزان مستقل بودن نتایج از داده‌های آموزش استفاده شد. برای این منظور، ابتدا داده‌ها به‌صورت تصادفی به پنج زیرمجموعه (5-fold CV) تقسیم‌بندی شدند. فرآیند مدل‌سازی پنج بار تکرار و در هر تکرار یک مجموعه جدید برای ارزیابی و چهار مجموعه دیگر برای آموزش مدل استفاده گشت. در پایان میانگین نتایج به‌عنوان دقت مدل گزارش شده است. با استفاده از روش اعتبارسنجی متقابل، هر یک از نمونه‌ها یک بار برای ارزیابی مدل استفاده می‌گردد. نتایج به‌دست آمده از طبقه‌بندی به کمک مدل آموزش دیده شده و با استفاده از داده‌های چندطیفی بیانگر میزان دقت طول‌موج‌های انتخاب شده برای جداسازی برگ‌های سالم و بیمار می‌باشد. از امتیاز F1 (F1-score) برای مقایسه دقت طبقه‌بندی استفاده شد (رابطه ۱).

$$F1 = \frac{2rp}{r+p} \quad (۱)$$

امتیاز F1 یک میانگین هارمونیک بین دقت (p) و یادآوری (r) می‌باشد. پارامتر دقت نسبت بین برگ‌های سالم صحیح طبقه‌بندی شده به کل برگ‌هایی که به‌عنوان برگ سالم توسط مدل طبقه‌بندی شده‌اند می‌باشد و پارامتر یادآوری نسبت بین برگ‌های سالم صحیح طبقه‌بندی شده توسط مدل به کل برگ‌های سالم است.

نتایج و بحث

همبستگی بین طول‌موج‌ها

شکل ۲ میزان همبستگی بین طول‌موج‌های مختلف در این مطالعه را نشان می‌دهد. همانطور که مشاهده می‌شود بین طول‌موج‌های فروسرخ (بیشتر از ۷۶۰ نانومتر) میزان همبستگی بالای ۹۵ درصد است. این مسئله دلالت بر افزونگی و اضافه بودن برخی از طول‌موج‌ها دارد. لذا این طول‌موج‌های اضافه باید حذف گردند. مثلاً اخذ یک طول‌موج در ناحیه قرمز باید کافی باشد زیرا طول‌موج‌های دیگر در این ناحیه اطلاعات جدیدی فراهم نمی‌کنند. بنابراین این طول‌موج‌ها فقط باعث افزایش ابعاد داده‌ها شده بدون افزودن هیچ اطلاعات جدید دیگری. شکل ۲ بر اهمیت انتخاب ویژگی تاکید می‌کند.

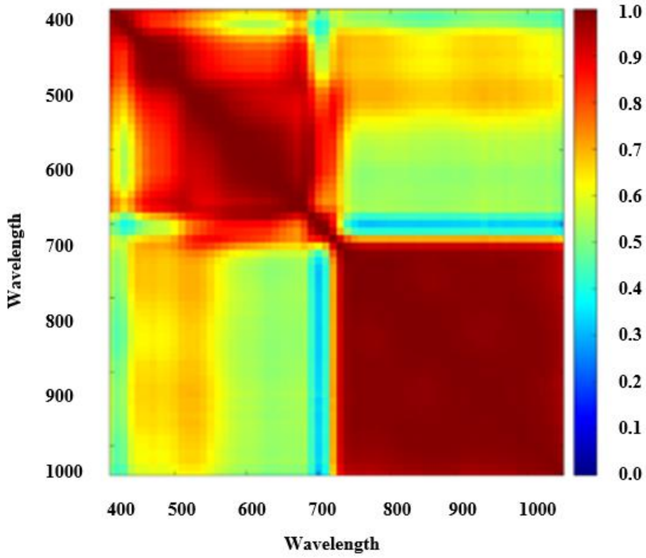
بازتابش برگ‌های سالم و بیمار

از بازتابش برگ‌های هر گروه (سالم و بیمار) میانگین‌گیری شد (شکل ۳). میزان میانگین بازتابش برگ‌های بیمار در ناحیه قرمز (۶۵۰ نانومتر) بیشتر از میانگین بازتابش برگ‌های سالم بود. بازتابش بیشتر برگ‌های بیمار در ناحیه قرمز نشان‌دهنده میزان کلروفیل کمتر در

1- Support Vector Machine (SVM)

برگ‌های بیمار می‌باشد زیرا در یک برگ سالم طول‌موج‌های قرمز (حدود ۶۵۰ نانومتر) و آبی (حدود ۴۵۰ نانومتر) توسط کروفیل برای انجام فتوسنتز جذب می‌شود (Al-Saddik *et al.*, 2018).

همچنین میانگین بازتابش برگ‌های بیمار در ناحیه فروسرخ به نسبت برگ‌های سالم کمتر شده است. بازتابش در ناحیه فروسرخ به ساختار سلولی گیاه ارتباط دارد و برگ‌های سالم بیشتر میزان نور تابیده شده در این ناحیه را بازتابش می‌کنند. بنابراین میزان بازتابش کمتر برگ‌های بیمار در مقایسه با برگ‌های سالم در این ناحیه گویای این مسئله می‌باشد که بیماری پسیل به ساختار سلولی برگ‌ها آسیب وارد کرده است که در نتیجه برگ تمایل به جذب بیشتری در این ناحیه دارد.



شکل ۲– میزان همبستگی بین طول‌موج‌ها در این مطالعه. همبستگی بالای ۹۵ درصد بین طول‌موج‌های فروسرخ وجود دارد.

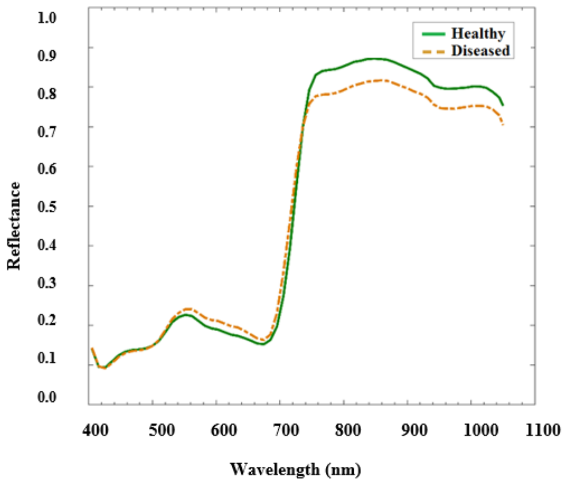
Fig. 2. Pairwise correlation between the spectral bands. The near infrared bands were highly correlated (>95%).

به‌علاوه طیف بازتابش‌یافته از روی برگ‌های بیمار در ناحیه لبه قرمز[1] امواج الکترومغناطیس که بین ناحیه قرمز (جذب بالا) و فروسرخ (بازتابش بالا) قرار دارد به سمت طول‌موج‌های کوتاه‌تر تمایل پیدا کرده است. این پدیده برای بیماری‌های دیگر و گیاهان دیگر نیز گزارش شده است (Mutanga and Skidmore, 2007; Sankaran *et al.*, 2011).

ترتیب طول‌موج‌ها بر اساس اهمیت

برای انتخاب ویژگی‌ها بر اساس اهمیت از الگوریتم جنگل‌های تصادفی استفاده شد. شکل ۴ میزان وزن هر یک از طول‌موج‌ها را نشان می‌دهد به‌طوری‌که این وزن‌ها نشان‌دهنده میزان اهمیت هریک از طول‌موج‌ها می‌باشد و مجموع وزن‌ها برابر با یک می‌باشد.

طول‌موج‌های ناحیه لبه قرمز بین ۶۹۵ تا ۷۱۵ نانومتر (به‌ترتیب ۷۱۵، ۷۰۵، ۷۲۵ و ۶۹۵ نانومتر) بیشترین اهمیت را داشتند. پس از ناحیه لبه قرمز، طول‌موج‌های فروسرخ حدوداً بین ۷۶۵ تا ۸۲۵ نانومتر دارای وزن و اهمیت بیشتری بودند. منحنی میزان اهمیت ویژگی‌ها دارای دو ماکزیمم محلی دیگر در ناحیه‌های بین ۶۳۵–۶۰۰ نانومتر و ۴۹۰–۴۶۰ نانومتر بود.



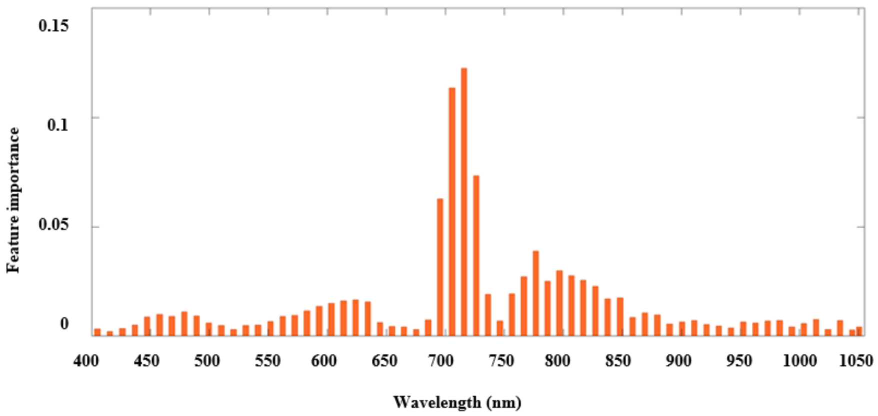
شکل ۳– میانگین بازتابش برگ‌های سالم و بیمار

Fig. 3. The average of spectral reflectance for the healthy and diseased leaves

نتایج خوشه‌بندی ویژگی‌ها

1- Red Edge

جدول ۱ نتایج خوشه‌بندی طول‌موج‌های دارای بیشترین وزن در شش خوشه و همچنین میانگین طول‌موج‌های هر خوشه را نشان می‌دهد.



شکل ٤– میزان اهمیت محاسبه شده هر یک از طول‌موج‌ها (ویژگی‌ها) توسط جنگل‌های تصادفی

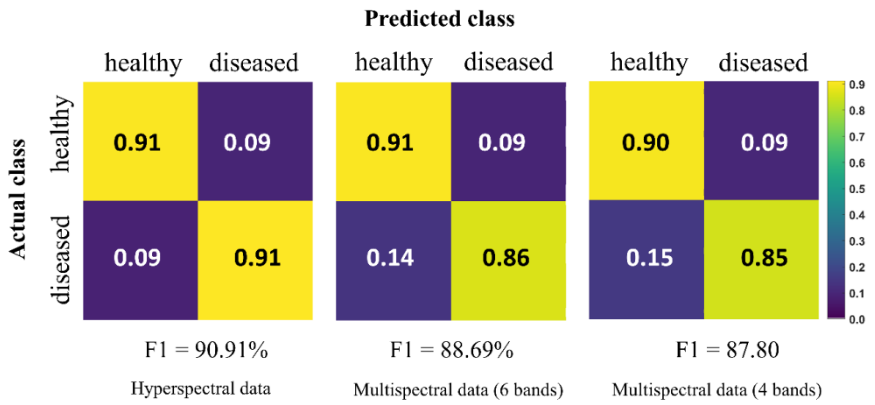
Fig. 4. Feature importance of the spectral bands calculated by random forests

جدول ۱– خوشه‌بندی طول‌موج‌های دارای بیشترین وزن در شش خوشه (طول‌موج‌ها برحسب نانومتر می‌باشند)

Table 1- Clustering of the spectral bands with the largest weights in six clusters (wavelengths are presented in nanometer)

	Cluster center	Members of each cluster							
Cluster 1	710	715^{1}	705^{2}	725^{3}	695^{4}	735^{13}	685^{33}		
Cluster 2	791	776^{5}	796^{6}	806^{7}	766^{8}	816^{9}	786^{10}	826^{11}	756^{12}
Cluster 3	858	847^{14}	837^{15}	868^{23}	879^{25}	858^{31}			
Cluster 4	598	623^{16}	613^{17}	634^{18}	603^{19}	592^{20}	582^{21}	572^{26}	561^{28}
Cluster 5	468	478^{22}	457^{24}	489^{27}	468^{29}	447^{30}			
Cluster 6	1023	1013^{32}	1034^{34}						

* Superscripts denote the rank of the features, ** Wavelengths are shown without decimal digits



شکل ۵– ماتریس آشفتگی به‌دست آمده برای طبقه‌بندی داده‌های فراطیفی و چندطیفی به کمک مدل ماشین بردار پشتیبان

Fig. 5. Confusion matrix obtained by support vector machine for classification of hyperspectral and multispectral data

با توجه با نمودار اهمیت طول‌موج‌ها و این نکته که میزان همبستگی بالایی بین طول‌موج‌های فروسرخ وجود داشت (شکل ۲) از میان مراکز خوشه‌ها، که در جدول ۱ نشان داده شده‌اند، فقط چهار طول‌موج ۷۱۰، ۷۹۱، ۵۹۸ و ۴۶۸ نانومتر انتخاب شدند.

نتایج طبقه‌بندی

پس از انتخاب مرکز باندهای چندطیفی، ماتریس داده‌ها از فراطیفی به چندطیفی تبدیل شد. شکل ۵ ماتریس آشفتگی برای طبقه‌بندی کلاس بیمار و سالم را نشان می‌دهد که توسط مدل ماشین بردار پشتیبان به‌دست آمده است.

همانطور که از شکل ۵ مشخص است تفاوت قابل ملاحظه‌ای بین امتیاز F1 طبقه‌بندی داده‌های فراطیفی و چندطیفی (۴ باند و ۶ باند) وجود ندارد. این مسئله دقت روش انتخاب ویژگی استفاده شده در این پژوهش را نشان می‌دهد. همچنین عدم تفاوت معنی‌دار بین دقت طبقه‌بندی با استفاده از تمام باندها و ۴ باند دلالت بر وجود همبستگی بالا بین باندهای فراطیفی و همچنین افزونگی در باندهای فراطیفی دارد.

نتیجه‌گیری

میزان اهمیت باندهای طیفی در ناحیه مرئی و فرو سرخ نزدیک (بین ۴۰۰ تا ۱۱۰۰ نانومتر) جهت شناسایی برگ‌های درخت پسته که آلوده به بیماری پسیل هستند به‌دست آمد. بر اساس میزان اهمیت ویژگی‌های طیفی و با استفاده از یک الگوریتم خوشه‌بندی ۶ طول‌موج به‌عنوان بهترین طول‌موج‌ها برای طبقه‌بندی برگ‌های سالم و بیمار پسته به‌دست آمد. سپس با میانگین‌گیری از طول‌موج‌هایی که در فاصله ۱۵ نانومتری این ۶ مرکز بود داده‌های فراطیفی (۶۴ باند) به چندطیفی (۶ باند) تبدیل شد. از آنجا که میزان همبستگی بین طول‌موج‌ها در ناحیه فروسرخ نزدیک بسیار بالا (بیش از ۹۵ درصد) بود از ۳ طول‌موج انتخاب شده در ناحیه فروسرخ نزدیک (۷۱۰، ۷۹۱، و ۱۰۲۳) فقط طول‌موج ۷۱۰ نانومتر که نزدیک‌تر به ناحیه مرئی بوده انتخاب شد. نتایج طبقه‌بندی برگ‌های آلوده و بیمار به استفاده از داده‌های فراطیفی و چندطیفی نشان داد که میزان دقت طبقه‌بندی در حدود ۲ درصد کاهش می‌یابد و اگر تنها از ۴ باند استفاده شود میزان دقت در حدود ۳ درصد کاهش می‌یابد.

نتایج این تحقیق برای انتخاب طول‌موج‌های مناسب می‌تواند در توسعه دو حوزه سخت‌افزاری (ساخت دوربین‌های چندطیفی) و نرم‌افزاری (کاهش ابعاد و تحلیل داده‌های فراطیفی) جهت تشخیص به هنگام بیماری پسیل و اتخاذ تصمیمات مدیریتی بسیار مفید و کاربردی باشد. به‌عنوان مثال با تجهیز پرنده‌های بدون سرنشین به دوربین‌های فراطیفی یا چندطیفی و استفاده از یافته‌های این تحقیق می‌توان نقشه دقیقی از توسعه و پراکندگی میزان آلودگی پسیل در باغ‌های استفاده کرد. با در اختیار داشتن نقشه آلودگی و با استفاده از تکنولوژی‌های نرخ متغیر (VRT) می‌توان از مزایای سم‌پاشی دقیق استفاده نمود. نتایج این تحقیق همچنین می‌تواند به پژوهشگران در انتخاب و توسعه شاخص‌های طیفی گیاهی جهت تحلیل تصاویر ماهواره‌ای فراطیفی و چندطیفی کمک نماید.

References

1. Al-Saddik, H., J. C. Simon, and F. Cointault. 2018. Assessment of the Optimal Spectral Bands for Designing a Sensor for Vineyard Disease Detection: The Case of Flavescence Dorée. Precision Agriculture (0123456789). doi.org/10.1007/s11119-018-9594-1.
2. Alisaac, E., J. Behmann, M. T. Kuska, H.W. Dehne, and A. K. Mahlein. 2018. Hyperspectral Quantification of Wheat Resistance to Fusarium Head Blight: Comparison of Two Fusarium Species. European Journal of Plant Pathology. doi.org/10.1007/s10658-018-1505-9.
3. Breiman, L. 2001. Random Forests. Machine Learning 45 (1): 5-32.
4. Hastie, T., R. Tibshirani, and J. Friedman. 2009. Springer Series in Statistics The Elements of Statistical Learning: Data Mining, Inference and Prediction. New York: Springer-Verlag New York. springerlink.com/index/10.1007/b94608.
5. Kheiralipour, K., H. Ahmadi, A. Rajabipour, Sh. Rafiee, M. Javan-Nikkhah, D. S. Jayas, and K. Siliveru. 2016. Detection of Fungal Infection in Pistachio Kernel by Long-Wave near-Infrared Hyperspectral Imaging Technique. Quality Assurance and Safety of Crops and Foods 8 (1): 129-35. doi.org/10.3920/QAS2015.0606.
6. Mahlein, A., E. Christian Oerke, U. Steiner, and H. Wilhelm Dehne. 2012. Recent Advances in Sensing Plant Diseases for Precision Crop Protection. European Journal of Plant Pathology 133 (1): 197-209.
7. Moghimi, A., C. Yang, M. E. Miller, F. Shahryar Kianian, and P. M. Marchetto. 2018. A Novel Approach to Assess Salt Stress Tolerance in Wheat Using Hyperspectral Imaging. Frontiers in Plant Science. doi.org/10.3389/fpls.2018.01182.
8. Moghimi, A., M. H. Aghkhani, A. Sazgarnia, and M. H. Abbaspour-Fard. 2011. Improvement of NIR Transmission Mode for Internal Quality Assessment of Fruit Using Different Orientations. Journal of Food Process Engineering 34 (5): 1759-74.
9. Moghimi, A., C. Yang, and J. A. Anderson. 2020. Aerial Hyperspectral Imagery and Deep Neural Networks for High-Throughput Yield Phenotyping in Wheat. Computers and Electronics in Agriculture 172: 105299. arxiv.org/abs/1906.09666.
10. Moghimi, A., C. Yang, J. A. Anderson, and S. K. Reynolds. 2019. Selecting Informative Spectral Bands Using Machine Learning Techniques to Detect Fusarium Head Blight in Wheat. In ASABE Annual International

Meeting, Boston, MA. doi.org/10.13031/aim.201900815. (August 13, 2019).
11. Moghimi, A., C. Yang, and P. M. Marchetto. 2018. Ensemble Feature Selection for Plant Phenotyping: A Journey from Hyperspectral to Multispectral Imaging. IEEE Access 6: 56870-84.
12. Mohammadigol, R., M. H. Khoshtaghaza, R. Malekfar, M. Mirabolfathi, and A. M. Nikbakht. 2013. Detection of Pistachio Aflatoxin Using Raman Spectroscopy and Artificial Neural Networks. Journal of Agricultural Machinery 5 (1): 1-9. (In Persian). http://dx.doi.org/10.22067/jam.v5i1.28122.
13. Mutanga, O., and A. K. Skidmore. 2007. Red Edge Shift and Biochemical Content in Grass Canopies. ISPRS Journal of Photogrammetry and Remote Sensing 62 (1): 34-42.
14. Nagasubramanian, K., S. Jones, S. Sarkar, A. K. Singh, A. Singh, and G. Baskar. 2018.Hyperspectral Band Selection Using Genetic Algorithm and Support Vector Machines for Early Identification of Charcoal Rot Disease in Soybean Stems. Plant Methods 14 (1): 1-13. doi.org/10.1186/s13007-018-0349-9.
15. Sankaran, S., A. Mishra, J. M. Maja, and R. Ehsani. 2011. Visible-near Infrared Spectroscopy for Detection of Huanglongbing in Citrus Orchards. Computers and Electronics in Agriculture 77 (2): 127-34. doi.org/10.1016/j.compag.2011.03.004.
16. Susic, N., U. Zibrat, S. Sirca, P. Strajnar, J. Razinger, M. Knapic, and B. Stare. 2018. Discrimination between Abiotic and Biotic Drought Stress in Tomatoes Using Hyperspectral Imaging. Sensors and Actuators, B: Chemical 273(June): 842-52.
17. Stefan, T., J. Behmann, A. Steier, T. Kraska, O. Muller, U. Rascher, and A. K. Mahlein. 2018.Quantitative Assessment of Disease Severity and Rating of Barley Cultivars Based on Hyperspectral Imaging in a Non-Invasive, Automated Phenotyping Platform. Plant Methods 14 (1): 45. doi.org/10.1186/s13007-018-0313-8.
18. Wahabzade, M., A. K. Mahlein, C. Bauckhage, U. Steiner, E. C. Oerke, and K. Kersting. 2016.Plant Phenotyping Using Probabilistic Topic Models: Uncovering the Hyperspectral Language of Plants. Scientific Reports 6(February): 22482. http://www.nature.com/articles/srep22482.